

A Bayesian Analysis of the Overattribution of Attitude

Donelson R. Forsyth and William R. Pope

Virginia Commonwealth University

Running Head: Overattribution

Date Submitted: May, 1980

PERMISSION TO REPRODUCE THIS  
MATERIAL HAS BEEN GRANTED BY

Wm. Pope

TO THE EDUCATIONAL RESOURCES  
INFORMATION CENTER (ERIC)."

Correspondence Address: Department of Psychology  
Virginia Commonwealth University  
810 W. Franklin Street  
Richmond, VA 23284

(804) 257-1179

Thanks are extended to Joel B. Cohen, Director of the Center for Consumer Research, University of Florida, for his help in many phases of this project. Requests for reprints should be addressed to the first author at the Department of Psychology, Virginia Commonwealth University, 810 W. Franklin Street, Richmond, VA 23284.



## Abstract

If overattribution--the tendency for observers to think actors' attitudes match their behaviors even when behaviors are not freely performed--results from the fundamental attribution error, then observers must believe the coerced behavior is attributionally informative. This assumption was tested by (1) investigating the extent to which attitude attributions made after observing the behavior differed from prior estimates of attitude and (2) assessing the diagnosticity of the behavior setting through examination of the Bayesian likelihood ratio. After measuring estimates of the distribution of favorable and unfavorable attitudes toward seatbelts, observers read either a pro- or anti-seatbelt essay that had supposedly been written under high or low choice conditions. Results supported the fundamental attribution error interpretation of overattribution since observers, rather than relying on prior estimates of attitudes and discounting coerced behavior, continued to assume that attitudes matched behavior even under low choice conditions.

### A Bayesian Analysis of the Overattribution of Attitudes

In a series of studies Jones and his colleagues (e.g., Jones, 1979; Jones & Berglas, 1976; Jones & Harris, 1967; Jones, Worchel, Goethals, & Grumet, 1971; Snyder & Jones, 1974) have confirmed predictions based on correspondent inference theory (Jones & Davis, 1965; Jones & McGillis, 1976) by finding observers tend to assume private attitudes match publicly expressed attitudes when environmental factors which could constrain behavior are minimal. However, this same research has also uncovered a pervasive tendency for observers to make attitude inferences even when environmental factors--and not internal attitudinal factors--are the more likely cause of expressed attitude. Jones (1979) has labelled this tendency to make inferences about an actor's attitudes on the basis of behavior performed in constraining situations overattribution, and has explained the effect in terms of the "fundamental attribution error:" Individuals mistakenly overestimate the information value of performed behaviors by maximizing the importance of personal dispositions while minimizing the role of environmental causes (Ross, 1977).

Although attitude attribution based on behaviors performed in constraining situations may appear to be produced by an attributional "error," Ajzen, Dalto, and Blyth (1979) have recently suggested that in some instances "overattribution" may result from relatively bias-free cognitive strategies. Several information processing approaches to attributions suggest that inferences concerning the attitudes of others are not only a function of the information contained in a single observation (e.g., one instance of writing

an essay), but also the subjective estimates of attitude based on past interactions. If, for example, the attributor observes an actor writing a pro-marijuana essay under highly restrictive conditions, then the attributor should logically discount the current behavior and fall back on prior estimates of attitudes. If the attributor has never met this individual and therefore has no target-based expectancies (cf. Jones et al., 1971), then category-based expectancies will be utilized to make an inference (Jones & McGillis, 1976). Thus, an attributor--in spite of the constraint of the situation--would still be justified in assuming the individual's attitudes match his or her behaviors if a pro-marijuana attitude is considered more probable based on estimates of the distribution of attitudes on the topic (e.g., Ajzen, 1977). Indeed, under conditions of high constraint an information integration approach to attitude attribution would predict that attribution of behavior-consistent attitude should increase as the attributor's subjective probability estimate of the likelihood of that attitude in the general population approaches unity (Trope, 1974; Trope & Burnstein, 1975).

This process of information integration is aptly described by Bayes' Theorem (Ajzen, 1971; Ajzen & Fishbein, 1975), which can be expressed as

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)} . \quad (1)$$

Interpreting equation (1),  $P(A|B)$  is the probability of a certain attitude given a certain behavior,  $P(B|A)$  is the probability of the behavior given the actor has a certain attitude,  $P(B)$  is the probability of the given behavior, and  $P(A)$  is the probability of the attitude. However, if it is assumed that attitudes can be either positive or negative on any particular

issue, then Bayes' Theorem is expressed as

$$\frac{P(A+|B)}{P(A-|B)} = \frac{P(A+)}{P(A-)} \times \frac{P(B|A+)}{P(B|A-)}, \quad (2)$$

or, when transformed to the logarithmic form recommended by Ajzen and Fishbein,

$$\log \frac{P(A+|B)}{P(A-|B)} = \log \frac{P(A+)}{P(A-)} + \log \frac{P(B|A+)}{P(B|A-)}. \quad (3)$$

In equation (3),  $P(A+|B)$  and  $P(A-|B)$  are the subjective conditional probabilities of a favorable or unfavorable attitude toward the topic discussed in the essay,  $P(A+)$  and  $P(A-)$  the prior subjective probabilities of a positive or negative attitude, and  $P(B|A+)$  and  $P(B|A-)$  are the subjective conditional probabilities of the behavior (writing an essay) given an individual has a positive or negative attitude. Thus,

$$\frac{P(A+)}{P(A-)}$$

are the subjective odds in favor of a positive attitude rather than a negative attitude (the priors),

$$\frac{P(A+|B)}{P(A-|B)}$$

are these same odds made after the essay (posteriors), and

$$\frac{P(B|A+)}{P(B|A-)}$$

reveals the degree to which the behavior indicates the actor has a positive rather than negative attitude. This final term, which is called the likelihood ratio (LR), summarizes the information value of the observed situation.

If, for example, the attributor decides that the actor was forced into writing the essay and hence it tells nothing about the writer's true

attitudes, then the probability of  $(B|A+)$  would be equal to the probability of  $(B|A-)$ . In other words, the actor would have made the essay irrespective of his or her attitude. If this were the case,  $\log LR$  would be 0.0, and (referring back to equation 3) attributions would be based solely on the remaining term in the formula: the prior probability estimates.<sup>1</sup> If, however, the attributor believes that the probability a person would write the essay is greater if the writer has a positive rather than negative attitude, then the likelihood ratio will be greater than zero and when combined with prior probability estimates will increase the posterior probability of positive attitudes and decrease the posterior probability of negative attitudes. The reverse should obtain if the attributor believes the probability a person would write the essay is greater if the writer has a negative rather than a positive attitude.

The present study applied Bayes' Theorem to the overattribution of attitudes to directly test the fundamental attribution error explanation proposed by Jones. If attributors--rather than simply discounting current behavior and relying on prior estimates of attitudes--are using behavioral information generated in highly restrictive settings to change their prior estimates of attitudes, then under low choice conditions posterior odds should differ from prior probabilities, or

$$\frac{P(A+|B)}{P(A-|B)} \neq \frac{P(A+)}{P(A-)} .$$

More specifically, if the actor writes a favorable essay, then Hypothesis 1 predicts

$$\frac{P(A+|B)}{P(A-|B)} > \frac{P(A+)}{P(A-)} .$$

Conversely, if the actor's speech or essay is unfavorable, then Hypothesis

2 is

$$\frac{P(A+|B)}{P(A-|B)} < \frac{P(A+)}{P(A-)} .$$

Second, if observers are revising their subjective probabilities on the basis of the behavior information, then they must believe the coerced behavior is attributionally informative. Therefore, under low choice conditions the logged likelihood ratio should be significantly different from 0.0, or

$$\frac{P(B|A+)}{P(B|A-)} \neq 0.0 .$$

More specifically, if the actor's positive speech is seen as informative, then Hypothesis 3 predicts

$$\frac{P(B|A+)}{P(B|A-)} > 0.0 .$$

If, on the other hand, the actor's negative speech is seen as informative, then Hypothesis 4 predicts

$$\frac{P(B|A+)}{P(B|A-)} < 0.0 .$$

Failure to find these effects would lend greater support to an information integration explanation of attitude attribution which argues "overattribution" does not result from an attributional "error" but from a rational tendency to rely on category-based information when target-based information is unavailable.

## Method

### Subjects

A total of 86 students--60 females and 26 males--were recruited from several undergraduate psychology courses for a study that was described as an investigation of how individuals formulate judgments of others on the basis of very limited information. The experimental materials were administered by a male experimenter in sessions involving 10 to 15 respondents, all of whom received course credit for their participation.

### Procedure

Each subject received a packet of materials which contained an informed consent sheet, several questionnaires, and an essay allegedly written by another student. The consent form explained that the materials in the packet involved several different studies, and included a survey of student attitudes on a range of issues, a correlational study of personality traits, and a study of how people form impressions of others.

Subjects first responded to the attitude survey, which was comprised of a series of statements concerning five contemporary issues. For each of the five topics, respondents were asked to indicate their degree of agreement with the stated position by placing an 'X' on a line which had nine segmented divisions between the end-points of "agree" and "disagree." Also, subjects estimated what percentage of students at the university would agree with the statement and what percentage would disagree with the statement. Embedded in the questionnaire was the statement "Seatbelts are



dangerous and should not be worn by drivers and passengers in automotive vehicles." This item provided both a measure of the subject's personal attitude as well as a measure of prior estimate of attitudes other students would have toward seatbelts  $[P(A+) \text{ and } P(A-)]$ .

The survey was followed by several personality scales which disguised the link between the attitude survey and the last questionnaire in the packet. This final set of materials, entitled the "Person Perception Study," was prefaced by directions based on Jones and Harris (1967). Supposedly, several students "here at the University" had provided the researchers with a variety of personal materials which might reveal something about their areas of expertise or interest. Subjects were to study the information provided--which could be a short autobiographical sketch, a completed attitude survey, grade transcripts, a formal resumé, or an essay on a current topic--and then record their perceptions.

In actuality, all subjects read an essay supposedly written by a student identified by the initials NLM. Before reading the essay subjects were told the essay writer had been free to choose the position argued in the essay (high choice) or that the position taken in the essay had been assigned (low choice). For approximately half of the subjects, a pro-seatbelts essay was included which argued that seatbelts are good and should be required in automobiles. The remaining subjects read an anti-seatbelts essay that took the position seatbelts are dangerous. Both essays were of equal length, and both used three arguments to support their position.

After reviewing the essay, subjects completed several questions concerning NLM. Two of these items served as checks of the choice and

side-taken-in-essay manipulations. Two others measured attitude attribution and confidence in attributions. Lastly, four items measured posterior inputs to the Bayesian equation, and included subjective probability estimates of the likelihood (1) that NLM is in favor of seatbelts  $[P(A+|B)]$ , (2) that NLM is opposed to seatbelts  $[P(A-|B)]$ , (3) that a person who favored seatbelts would write an essay like NLM's  $[P(B|A+)]$ , and (4) that a person who opposed seatbelts would write an essay like NLM's  $[P(B|A-)]$ .

### Results

Subjects' responses were examined in 2 (high vs. low choice) by 2 (pro-seatbelt essay vs. anti-seatbelt essay) analyses of variance which used a least-squares procedure to adjust each effect for other effects of equal or lower order. Following previous analyses (e.g., Ajzen, 1971; Trope, 1974), (1) the logarithm of the attitudinal and behavioral probability estimates was used in all analyses to avoid extreme scores and promote ease of interpretation, and (2) probability estimates equal to 0 were set to .01 since  $\log_2 0$  is undefined. In addition, because initial analyses which included sex of respondent as a classification variable yielded no differences between males and females, this factor was not retained in subsequent analyses.

#### Manipulation Checks

The choice manipulation proved to be effective. All respondents were asked to indicate how much choice NLM had "in selecting which side of the issue to argue for" using a 9-point scale which ranged from "no choice at all" to "very much choice." Only a main effect of choice,  $F(1, 80) = 30.16$ ,  $p < .05$  reached significance on this item, indicating low choice respondents

attributed less freedom to NLM than did respondents in the high choice condition; the means were 3.3 and 6.6, respectively. In addition, all but two subjects correctly indicated which type of essay NLM had written. The two female respondents who mistakenly identified NLM's anti-seatbelt essay as an essay favoring seatbelts were deleted from subsequent analyses.

#### Subjective Probability Estimates

Prior probabilities. Before exposure to the essay, subjects recorded their estimates of student seatbelt attitudes. Overall, subjects indicated that 50.5% of the student population would be pro-seatbelts, while approximately 49.5% would be anti-seatbelts. Hence, the log of prior estimates of positive versus negative attitude toward seatbelts  $[P(A+)/P(A-)]$  was +0.17.

Posterior probabilities. Basing computations on equation (3), the ratio of the probability of a favorable attitude given the behavior  $[P(A+|B)]$  to the probability of an unfavorable attitude given the behavior  $[P(A-|B)]$  was calculated to represent posterior attitude estimates. Because the data are based on logarithms, a posterior probability of 0.0 indicates either attitude was considered equally likely, a negative probability indicates an anti-seatbelt attitude was considered more likely than a pro-seatbelt attitude, and a positive probability indicates a pro-seatbelt was considered more likely than an anti-seatbelt attitude.

Analysis of this quotient revealed a significant interaction of essay direction and choice;  $F(1, 79) = 5.81, p < .05$ . As the means in Table 1 indicate, when NLM could choose which side to argue for, an anti-essay

made a negative attitude seem much more likely than a positive attitude, while a pro-essay made a positive attitude seem more likely than a negative attitude. However, even if the essays were written under low choice conditions, students who wrote pro-essays were viewed as more in favor of seatbelts than students who wrote anti-essays. This overattribution in the low-choice/pro-essay condition confirmed Hypothesis 1 since the posterior probability of a favorable attitude, +3.95, was significantly greater than the prior estimate of a favorable attitude; +0.17. The one-tailed  $t$  (38) was 3.82,  $p < .05$ . In addition, although the anti-essay written under low choice conditions made an unfavorable opinion seem slightly more likely than a favorable opinion, the change from a prior probability of +0.17 to a posterior probability of -0.54 was not statistically significant;  $t$  (48) = -0.84,  $p < .20$ . Hence, Hypothesis 2 received only directional support.

-----  
 Insert Table 1 about here  
 -----

### Information Diagnosticity

The interaction of choice and essay direction was also significant for the log of the likelihood ratio  $[P(B|A+)/P(B|A-)]$ ;  $F(1, 80) = 11.89$ ,  $p < .05$ . As the means presented in Table 1 indicate, under high choice conditions, a pro-essay made a positive attitude toward seatbelts seem more likely while an anti-essay made a negative attitude seem more likely. However, these same attributional tendencies were found under low choice conditions, thus lending support to the fundamental attribution error explanation of overattribution. Supporting Hypothesis 3, a pro-essay written under high constraint was still taken as evidence of a positive attitude

toward seatbelts; the likelihood ratio of 2.4 was significantly different from zero (the correct comparison since 0.0 equals 1.0 when dealing with logged data);  $t(38) = 2.71$ ,  $p < .05$ . Hypothesis 4 was also confirmed since an anti-essay written under high constraint was still taken as evidence of a negative attitude toward seatbelts; the likelihood ratio of -3.5 was also significantly different from zero;  $t(48) = -3.40$ ,  $p < .05$ .

### Ancillary Findings

Attitude attribution. Subjects not only reported attitude attribution through probability estimates, but also by responding to a more traditional 9-point Likert-type item which queried "In your estimation, what do you feel NLM's true attitude about seatbelts is?" The choice X essay direction interaction was significant,  $F(1, 80) = 14.85$ ,  $p < .05$ , and furnished additional evidence of overattribution. As the means in Table 1 show, attributed attitude matched essay direction under high choice conditions, just as correspondent inference theory would predict (e.g., Jones, 1979). Attitude attribution still occurred, however, under the low choice conditions, particularly when the essay favored seatbelts. Hence, the pattern of the means closely matches the same pattern found for subjective estimates of posterior probability. Indeed, the correlation between  $[P(A+|B)/P(A-|B)]$  and responses to this item was  $-0.80$  ( $n = 83$ ),  $p < .05$ .

Bayesian predictions. Based on prior probabilities and likelihood ratios, the amount of attitude attribution predicted by Bayes' Theorem was calculated using equation (3). Once more, the interaction of choice and essay direction proved significant;  $F(1, 78) = 5.67$ ,  $p < .05$ . The means

are presented in the final row of Table 1, and they show that the Theorem predicted a negative attitude was more likely than a positive attitude when an anti-seatbelt essay had been written under either high or low choice conditions ( $p_s < .05$ ). Conversely, Bayes' Theorem predicted a positive attitude was more likely than a negative attitude when the essay argued in favor of seatbelts and had been written under high choice conditions ( $p < .05$ ). Lastly, although the effect was somewhat reduced if NLM wrote the pro-essay under no choice conditions, a positive attitude was still predicted by the Theorem since the prediction ratio of 1.86 is significantly different from 0.0 ( $p < .05$ ). As would be expected based on the similar patterns obtained on all the attribution measures, the Bayesian prediction of attitude correlated significantly ( $p_s < .05$ ) with posterior probability estimates ( $r = +.48$ ,  $n = 81$ ) and the more traditional Likert-type measure of attribution ( $r = +.50$ ,  $n = 82$ ).

Confidence and own attitudes. Subjects reported their confidence in making judgments about NLM's attitude and in their estimate of posterior probabilities on a 9-point scale ranging from "not confident" to "very confident." No between condition differences were found on this item, as most subjects expressed a relatively high degree of confidence; the mean was 6.8. Also, while the pre-measure of subjects' personal attitudes toward seatbelts indicated most favored them ( $M = 7.4$ ), attitude did not correlate with any of the other dependent measures.

### Discussion

These results indicate that attitude attributions identified by previous research as overattributions do apparently stem from an

information integration bias in the inferential process. By measuring the probability the actor possesses a particular attitude both before and after exposure to the behavioral information, the impact of the new information on attributions was considerably clarified. Although attributors did not consider a positive attitude to be more likely than a negative attitude before reading the essay, after exposure to the essay attributions shifted in the direction of the position taken in the essay. Although this shift is quite reasonable when the writer's choice of essay direction was not constrained, the shift becomes overattribution when the essay was written under high constraint. Supporting the overattribution hypothesis, when a pro-essay was written under low choice conditions, the posterior odds in favor of a positive attitude were significantly greater than the prior odds in favor of the positive attitude. In addition, when an anti-essay was written under low choice conditions, the posterior odds in favor of a negative attitude became somewhat smaller than prior odds favoring a negative attitude, although the shift was not significant.

Examination of the likelihood ratio from the Bayesian analysis, which is an indicator of the assumed diagnosticity of the information, also indicated that the overattribution in the low choice conditions was produced by observers' assumptions that the behavioral information did say something about the actor's attitudes. As a fundamental attribution error interpretation of overattribution would predict, observers assumed that (1) a person would be more likely to write a pro-essay if they had a positive attitude rather than a negative one, and (2) a person would be more likely to write

an anti-essay if they had a negative attitude rather than a positive one--even when the link between attitude and essay was broken by situational constraints.

In terms of both posterior probability estimates and likelihood ratio, the anti-essay written under low choice conditions was viewed as somewhat less informative than the pro-essay. In other words, overattribution was less pronounced in the anti-essay/low choice condition. One possible explanation for this difference between essays focuses on the different expectations concerning the likelihood of each position. As Jones (1979) has recently pointed out, overattribution seems to occur more frequently when the position being advocated is expected rather than unexpected. In the current experiment, it was the case that a pro-seatbelt attitude was viewed as more likely than an anti-seatbelt attitude, but the difference in expectations--at least as assessed by prior probabilities--was quite small. However, the expectations explanation may still explain the observed pattern of results if respondents, despite their indication of equal likelihoods, still "expected" a positive attitude. Most of the subjects in the current investigation favored seatbelts, and a "false consensus bias" may have led them to expect the essay writer would also be pro-seatbelts (Ross, 1977). Hence, subjective probability did not translate directly into subjective expectations.

Although the application of Bayesian principles to the attitude-attribution paradigm yielded strong support for the fundamental attribution error explanation of the overattribution effect, the current investigation agrees with Ajzen et al. (1979) that it may be more accurate



to ascribe overattribution to a "cognitive strategy" rather than an "attributional error." Subjects in this study formulated their attributions in a relatively logical manner by combining information extracted from the observed situation with prior information about the distribution of attitudes among university students. Indeed, their attributions were fairly well predicted by a Bayesian equation which made use of subjects' prior estimates of attitude and likelihood of behavior ratios. Thus, while subjects did "overattribute" attitude, their attributions were the outgrowth of a logical cognitive strategy rather than a gross error in judgment. The mistake did not occur as subjects were formulating impressions of the actor's attitudes, but rather when they were appraising the strength of the constraining factors present in the situation. This result serves as a reminder that the fundamental attributional bias documented by Ross (1977) is dualistic in nature; it results not only from the perceptually prominent link between the actor and the act, but also from the weakness of the link between the situation and the act. Thus overattribution does not occur only because behavior engulfs the perceptual field, but also because situational factors are often lost in that field.

## References

- Ajzen, I. Attribution of disposition to an actor: Effects of perceived decision freedom and behavioral utilities. Journal of Personality and Psychology, 1971, 18, 144-156.
- Ajzen, I. Intuitive theories of events and the effects of base rate information on prediction. Journal of Personality and Social Psychology, 1977, 35, 303-314.
- Ajzen, I., Dalto, C., & Blyth, D. Consistency and bias in the attribution of attitudes. Journal of Personality and Social Psychology, 1979, 37, 1871-1876.
- Ajzen, I., & Fishbein, M. A. Bayesian analysis of attribution processes. Psychological Bulletin, 1975, 82, 261-277.
- Jones, E. E. The rocky road from acts to dispositions. American Psychologist, 1979, 34, 107-117.
- Jones, E. E., & Berglas, S. A recency effect in attitude attribution. Journal of Personality, 1976, 44, 433-448.
- Jones, E. E., & Davis, K. E. From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 2). NY: Academic Press, 1965.
- Jones, E. E., & Harris, V. A. The attribution of attitudes. Journal of Experimental Social Psychology, 1967, 3, 1-24.
- Jones, E. E., & McGillis, D. Correspondent inference and the attribution cube: A comparative reappraisal. In J. H. Harvey, W. J. Ickes, & R. F. Kidd (Eds.), New directions in attribution research (Vol. 1). Hillsdale, NJ: Erlbaum, 1976.

- Jones, E. E., Worchel, S., Goethals, G. R., & Grumet, J. F. Prior expectancy and behavioral extremity as determinants of attitude attribution. Journal of Experimental Social Psychology, 1971, 7, 59-80.
- Ross, L. The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 10). NY: Academic Press, 1977.
- Snyder, M., & Jones, E. E. Attitude attribution when behavior is constrained. Journal of Experimental Social Psychology, 1974, 10, 585-600.
- Trope, Y. Inferential processes in the forced compliance situation: A Bayesian analysis. Journal of Experimental Social Psychology, 1974, 10, 1-16.
- Trope, Y., & Burnstein, E. Processing the information contained in another's behavior. Journal of Experimental Social Psychology, 1975, 11, 439-458.

## Footnote

<sup>1</sup>Log<sub>2</sub> of 1.0 = 0.0.

Table 1

Posterior Odds, Likelihood Ratio,  
Observer Attributions, and Bayesian Predictions

Item	Anti-essay		Pro-essay	
	Low Choice ( <u>n</u> = 20)	High Choice ( <u>n</u> = 19)	Low Choice ( <u>n</u> = 25)	High Choice ( <u>n</u> = 20)
Posterior Odds for Positive Attitude	-0.54 <sub>b</sub>	-4.67 <sub>a</sub>	3.95 <sub>c</sub>	3.88 <sub>c</sub>
Likelihood Ratio for Positive Attitude	-3.46 <sub>b</sub>	-8.50 <sub>a</sub>	2.40 <sub>c</sub>	4.48 <sub>c</sub>
Observers' Attribution of Attitude	3.44 <sub>b</sub>	1.15 <sub>a</sub>	5.85 <sub>c</sub>	6.53 <sub>c</sub>
Bayesian Prediction of Attitude	-4.03 <sub>a</sub>	-7.68 <sub>a</sub>	1.86 <sub>b</sub>	5.51 <sub>b</sub>

Note. Cell ns are presented in parentheses. The more positive the mean, the greater the attribution of a favorable attitude, except for the likelihood ratio where an increase in absolute value implies an increase in informational value. Means without a single common subscript differ at  $p < .05$ .